Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems

Sammy V. Militante
College of Engineering and Architecture
University of Antique
Sibalom, Antique, Philippines
https://orcid.org/0000-0003-1962-9450

Nanette V. Dionisio
College of Arts and Sciences
University of Antique
Sibalom, Antique, Philippines
nanette_dionisio@yahoo.com

***Abstract*— COVID-19 or Severe Acute Respiratory Syndrome Corona virus-2 is an extremely transmissible virus that is discharged through breathing droplets released from an infected individual who is talking, sneezing, or coughing. Close interaction with a person infected or through touching a contaminated surface and object can spread the virus rapidly. As of now, there is no vaccine to combat the COVID-19, and the best way to protect the person from a virus is to avoid being exposed to it. Wearing a facemask that covers the nose and mouth in a public setting and repeatedly cleansing of hands or the use of at least 70% alcohol-based disinfectants is a practice to avoid virus exposure. Deep Learning technology has demonstrated its achievement in recognition and classification by processing images. The research study uses deep learning techniques that identify if the person is wearing a facemask or not and check if the persons in the area observe physical distancing. The collected image data contains 20,000 images, uniformly crop images in 224x224 pixels, and attained an accuracy rate of 97% during the training of the model. The developed system is implemented using Python and OpenCV through TensorFlow that recognizes persons wearing a facemask or not wearing and measures the physical distance between each person. It signals an alarm and captures facial images upon detecting persons not wearing a mask and does not observe physical distancing. This study is beneficial in combating the spread of the virus and avoiding contact with the virus.**

Keywords—Facemask Detection, Physical Distancing Detection, Alarm System, COVID-19, Deep Learning

I. INTRODUCTION

As the Coronavirus outbreak continues, most of the people in the world are suffering badly due to this pandemic. Every day thousands of people are dying globally because of the COVID-19 virus. The World Health Organization Coronavirus (COVID-19) situation report as of August 8, 2020, there are over 19 million people infected with COVID-19 disease across 213 countries in the world and it killed over 716 thousand people globally [1]. The Philippines is now on the top list of most cases reported with the infected Coronavirus in the Southeast Asian (ASEAN) Region [2]. It has reported that it has 126,885 infected cases and it has claimed 2,209 lives of the Filipinos [1-2]. Coronavirus disease 19 or COVID-19 is a respiratory illness that caused severe pneumonia in an infected person [3] and acquiring this disease is through person-to-person direct contact with generated respiratory droplets, droplets of saliva or discharge from the nose when the infected person coughs or sneezes or through breathing in the virus if you are within the range of an infected person. You can also acquire this virus by touching a contaminated surface and then entering the virus in your mouth, nose, or eyes [4]. Elderly persons, persons with a poor immune system, and those with underlying health complications like heart failure, diabetes, severe respiratory disease, and cancer are vulnerable in this type of illness [5].

At this time, there is no cure, or no vaccines developed to treat the COVID-19 disease but there are many on-going clinical trial evaluations for the possible treatment of this disease [4]. The spread or transmission of coronavirus can be prevented by observing protocols set by medical authorities. The best way to prevent and to slow-down the further spread is to protect yourself and others by washing your hands regularly or using a disinfectant and alcohol with 70% solutions [3] and by not-touching your face including nose, eyes, and mouth [6]. The virus spread can be limit by observing social distancing and observing hygiene like compulsory wearing of facemask, use of hand-gloves, face-shield, and the use of sanitizer is very important [3]. The Government and most organizations are making it compulsory to follow social distancing and the wearing of a facemask. Facemask detection system through artificial intelligence is a ground-breaking technological answer that is beneficial to everyone is recognizing people wearing facemasks [7]. A new normal has emerged in this pandemic by covering faces with a mask that is very effective in the prevention of the disease outbreak. On the other hand, it will be challenging to recognize faces with masks on any monitoring systems while maintaining contactless access control in buildings or premises [8-9]. Covering faces with a mask is a challenge posed for face detection algorithms and performance [9]. Physical distancing is an exercise of staying at a protected distance of at least 2 meters from the other person to avoid being contracted with the virus [8]. Observing these precautionary measures will prevent being contaminated with the coronavirus.

Advancements in the field of deep learning, particularly convolutional neural networks (CNNs), have already shown remarkable success in the classification of images [10]. The key idea behind the CNNs is to create an artificial model, like a visualization area of the human brain [11]. The biggest advantage of CNNs is that one can extract more important characteristics over the whole image, instead of just handcrafted attributes [10]. Researchers introduced different deep networks based on CNN, and these networks achieved the state of results in computer vision classification, segmentation, object detection, and localization.

In this research study, deep learning techniques are applied to construct a classifier to collect images of a person wearing a face mask and not wearing from the database, it can distinguish between those classes of facemask wearing and not facemask wearing, and measures the distance between people that observes physical distancing then creates an alarm if it is not observed properly including the wearing of the facemask. The artificial neural network has proven to be a vigorous process for extracting features from unprocessed data. This study proposes the use of a convolutional neural network to design the facemask classifier and to include the

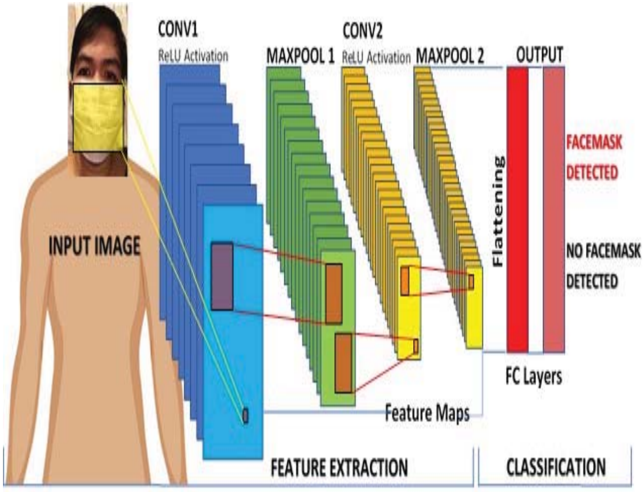

Fig. 1. Process of Convolutional Neural Network in predicting Facemask and not Facemask wearing.

effect on the predictive performance of the number of the convolution neural layer.

II. Related Works

A. The CNNs or Convolutional Neural Networks

CNN's is a conventional multi-layer neural network where the previous layer feeds one layer and outcomes can be measured and analyzed from both layers [12]. CNN is applied precisely in image-processing [10], processing of human-language, computer-vision [11], self-driving automobiles. The CNNs is also a particular form of neural network architecture that depicts a conventional feed-forward neural network; it simulates a human visual processing cortex of the brain, where the filtration system is a network of cells that resembles a certain portion of a picture [7]. A convolutional-layer in a neural network consists of a filter that converts over the input according to the particular size of the filter and the stride width [5]. The illustration in figure 1 provides an example of the convolutional-layer process. The filter will model completely the image elements in the filter and map it to a new array, then the filter will stride to the next location and iteratively perform the same procedure.

Feature maps are created as the filter travels through the image, projecting that very same pixel into the output layer multiple times [13]. The output layers distribute the same parameterization which produces shared weights. The shared weights correlate to the property whereby similar objects can be identified across an image regardless of the actual position. The pooling-layer follows immediately after a convolutional layer integrates multiple artificial outputs into a single neuron, reducing the weight of the layers behind [5]. For instance, in max pooling, the pooling function selects the elements with the largest value to be mapped to the output neuron while in average pooling the function averages the elements and the average of the elements in the filter is selected instead.

B. The Deep Learning Techniques

Deep learning is essentially a combination of artificial intelligence and machine learning. Inspired by brain neurons, this has proven greater flexibility and builds more accurate models compared to machine learning. Deep learning has applications in many fields like image classification, speech recognition, computer vision, natural language processing, bioinformatics, advertising, e-commerce, digital marketing, robot learning, and many others.

Fig. 2. Shows photographic images of a person wearing a facemask and not wearing a facemask dataset.

III. Proposed Methodology

A. Facemask and No-Facemask Dataset Collection

The collection of images used for training and testing the model was collected from the web. A python script was developed that uses a Bing Search APIs that create a web associated applications and services that will search webpages, images, news, videos, spell-check, etc. and download these images automatically to the directory with appropriate labeling from the web. Other images were taken specifically in the University of Antique with selected students and personnel as part of data collection. The dataset was able to generate 20,000 images using 2 classes, facemask wearing, and no-facemask wearing. The image size resolution of the dataset ranges from 800 up to 1200 pixels. All images were formatted in a JPEG formatting. The sample of the collected image dataset is illustrated in figure 2 illustrating the person wearing a facemask and not wearing a facemask.

B. Activation

There are multiple activation functions and it does not yield similar outcomes due to the various statistical architecture. Soft-max can be a preferred alternative if there is a multi-label classification problem, and this should be avoided for binary classification [14]. Leaky ReLU can be used when most of the input value emerges as 0, as the ReLU function releases more dead neurons. Dead neurons could not help making choices. ReLU is the most widely used activation feature because of lower computational costs and equivalent accessibility in hidden layers of the network [15].

C. Transfer Learning and Fine-Tuning

Fine-tuning is the final method in the training of the model and the start-off point of transfer learning in the study. Feature extraction is used in the first layers of each CNN architectures and the last few layers are intended for learning. Each architecture learned two classes of learned facemask wearing and not facemask wearing and the last few layers of every architecture are replaced with the same kind of layers but with different parameters. The architecture is trained and fine-tuned with the initial settings set by the authors [16]. The model is designed to learn only two classes of facemask wearing and not wearing with the weight rate and neuron bias rate of 20 on both the fully connected layers and SoftMax or activation to accelerate the processing of learning on new layers.

D. Classification

Classification is done with the training of the model that will recognize and classify the trained images with the learned visual patterns [17]. The authors implemented the development of a program using open-source software using Python and OpenCV. Along with the TensorFlow and Keras module to create a CNN model using the VGG-16 network model. The study applies a supervised learning model with 80% of datasets were used in training while 15% is used for testing and 5% for validation. The input parameters for the input image are set to 224 for its height and width, the batch sizes of 32 and 100 iterations. Subsequently, the method of data augmentations is also utilized for image data build-up by basically employing rotation, rescaling, shifting, and zooming procedures. The dropout rate is set at 50% and the rescaling of 1/255 which leads to a multiple for each pixel image. The horizontal flip, nearest fill-mode, 0.3 factor for zoom range, 0.2 width shift range, horizontal and vertical shift factors, and 20 rotation range were utilized as parameter settings.

IV. Experimental Setup

The entire simulation was performed in the 32-bit Windows10 Operating System, using the application development language version of Python 3.6. The experiment was coded to develop and train the model using Keras as backend and within the Tensor-flow platform. Also used to implement the system is the laptop computer with Intel ® Core™ i7-8700 CPU @4.60GHz with 12 M Cache, 16 GB RAM, and GTX 1050 video card. Data augmentation was also incorporated to increase the image data collected during the conduct of the research. Some of these techniques used are image-flipping, image-rotation, zoom-range, and range-shifts. By futzing with the weights, the optimizer designs and shapes the model to the most realistic achievable form.

Optimizers are methods that changes the attributes of the neural networks like weights and learning-rate for reducing the losses during training. As features of optimization were introduced in the analysis, Adam optimizer with a learning rate of 0.0001, and categorical cross-entropy. Adam optimizer is the combination AdaGrad and RMSProp algorithms that can control sparse-gradients on noise-problems. Validity accuracy diagrams relative to consistency and loss of validity as opposed to a loss of training are shown in Figure 3. Two hidden layers are pretty important in a deep learning model. Combining the output of one or more hidden layers as one output layer is important. By using more hidden layers, it offers a deeper analytical model on the one hand but on the other hand, each added layer adds complexity in computation.

Besides, higher numbers of neurons added in each layer will also result in high computational costs. Some added features to the CNN model are optimization using Adam and categorical cross-entropy is implemented. Adam optimizer has shown good results in the application of computer vision and natural language processing. Adam's design structure is pretty attractive. For the entire neural network and each attribute, the learning weights and descent rate are the same. On the other hand, Adam provides different learning levels for various parameters and heightens the model's overall efficiency. Adam sets new weights for increasing time-based on a previous value of an attribute. Two types of cross entropy are useful in the problem of profound learning classification. The loss function generally tells the discrepancy which is nothing but the error between the actual output and the predicted output. Categorical cross-entropy is useful when clustering data instances among multi-class labels are needed. Epoch implies a full list of available inputs. As when weights are calculated for developing a model after each epoch. The weights change and again tested against the next cycle simulation of the same dataset (called next epoch). The entire training data is expected to be in the main memory when executing this. For larger datasets, keeping the whole dataset in the main memory is not feasible at multiple times, so the whole epoch (dataset) is divided into batches and each batch is sequentially brought into main memory and executed, and the result is summed up and finally interpreted as an epoch output.

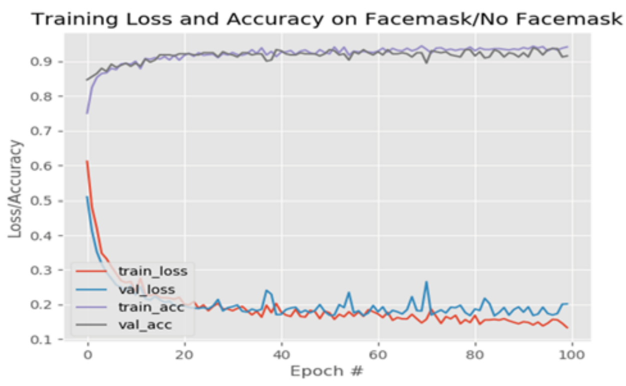

Fig. 3. During model training the result of the Accuracy / Loss performance test.

An alarm system is integrated in the system using pyttsx3 a text-to-speech platform that produces an alarm and a voice notification that prompts a person(s) not wearing a facemask and do not observe physical distancing. The system is also designed to count the number of persons violating the observance of physical distancing and it uses the real-time from the system to display the time it captures the person wearing a facemask and not-wearing facemask.

V. Result and Discussions

Very acceptable validation accuracy was achieved during the training of the CNN model through several experiments conducted and it has a recorded rate of 96% with batch sizes fixed to 32 and 100 iterations for epochs as illustrated in fig. 3 for performance tests results from visualization through accuracy and loss. Figures 4 and 5 display the test results on the performance of the model in detecting persons wearing a facemask with a rate of 97.87% and 98.75% detected facemask wearing respectively. Analyses of the outcome of test images with a rate of 98.20% on the left picture of figure 6 and 98.82% on the right picture of figure 6 show that it has detected with no facemask wearing and a 95.15% accuracy rate on figure 7 displays multiple people detected with no facemask wearing. Figure 8 describes the detection of four persons violating the social distancing rule that is marked on a rectangular frame with a red color while a rectangular frame mark with green color are those people observing the proper social distancing. The same detection is illustrated in figure 9 and figures 10 with two persons and six persons violating the social distancing mark in a color red rectangular frame correspondingly. Figures 11 and 12 illustrate the persons in the picture observing the social distancing protocol mark in a color green rectangular frame.

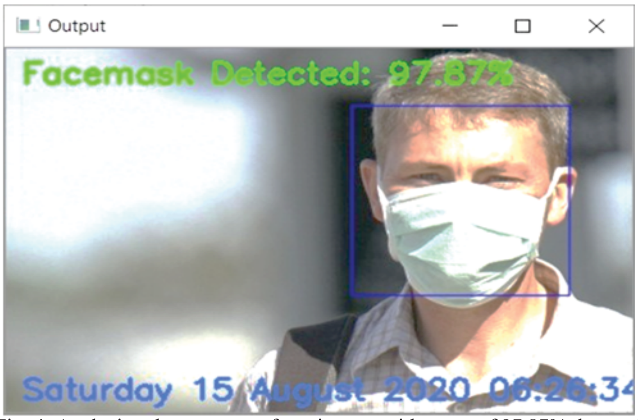

Fig. 4. Analyzing the outcome of test images with a rate of 97.87% detected facemask wearing.

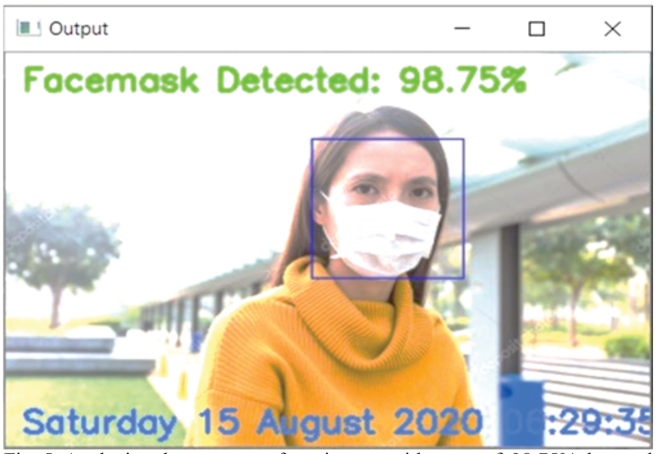

Fig. 5. Analyzing the outcome of test images with a rate of 98.75% detected facemask wearing.

Fig. 6. Analyzing the outcome of test images with a rate of 98.20% on the left picture and 98.82% on the right picture that it has detected with no facemask wearing.

Fig. 7. Analyzing the outcome of test images with a rate of 95.15% that it has detected with no facemask wearing.

Fig. 8. The output result of a picture with four persons violating the social distancing mark on the color red.

Fig. 9. The output result of a picture with two persons violating the social distancing mark in the color red.

Fig. 10. The output result of a picture with six persons violating the social distancing mark in the color red.

Fig. 11. The output result of a picture with persons observing the social distancing mark with the green color.

Fig.12. The output result of a picture with persons observing the social distancing mark with the green color.

VI. Conclusions and Recommendations

This paper manuscript presents a solution in preventing the transmission of the COVID-19 pandemic. The study focuses on detecting the face mask-wearing and facemask not wearing in real-time with an alarm system and also detects the observation of social distancing between persons walking in the community using the application of deep learning techniques through Convolutional Neural Networks. The authors were able to develop a VGG-16 model that gives precise and speedily results for facemask detection. The trained model was able to generate a result of 97%. The test result shows a notable accuracy rate in detecting persons wearing a facemask and not wearing a facemask and also detects the person if they are observing the physical distancing or not. Likewise, the study presents a useful tool in fighting the spread of the COVID-19 virus by detecting a person who wears a facemask or not, calculates the distance between each person if they are observing physical distancing or not, and sets an alarm if these rules are not observed.

Future works include the integration of face shield detection, in which the camera detects the person wearing a facemask or not, at the same time measures the distance between each person, check and detect the person wearing a face shield or not as additional protection in COVID-19 prevention and creates an alarm if these rules are not observed properly. The combination of several models of CNNs and evaluate every model with the highest performance accuracy during training to increase the performance in detecting and recognizing people wearing facemasks and face shield is suggested.

Acknowledgment

The authors express their gratefulness to the University of Antique for carrying out this research project.

References

[1] World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report– 201, 2020, [cited 8 August 2020], https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200808-covid-19-sitrep-201cleared.pdf?sfvrsn=121bb855_2

[2] Department of Health: Updates on Novel Coronavirus Disease (COVID-19), 8 August 2020, https://www.doh.gov.ph/2019-nCoV

[3] S. Militante, N. Dionisio, "Real-Time Facemask Recognition with Alarm System using Deep Learning", Paper presented at the 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC 2020), 8 August 2020, Shah Alam, Malaysia

[4] Jefferson, T. et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. Cochrane Database Syst Rev, CD006207, https://doi:10.1002/14651858.CD006207.pub4 (2011).

[5] S. Militante, N. Dionisio, B. Sibbaluca, "Pneumonia Detection through Adaptive Deep Learning Models of Convolutional Neural Networks," Paper presented at the 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC 2020), 8 August 2020, Shah Alam, Malaysia

[6] Feng, S., Shen, C., Xia, N., Song, W., Fan, M., & Cowling, B. J. (2020). Rational use of face masks in the COVID-19 pandemic. The Lancet. Respiratory medicine, 8(5), 434–436. https://doi.org/10.1016/S2213-2600(20)30134-X

[7] S. V. Militante, and B. G. Sibbaluca, "Pneumonia Detection Using Convolutional Neural Networks", International Journal of Scientific & Technology Research, Volume 9, Issue 04, pp. 1332-1337, April 2020.

[8] Zhang, K. P., Zhang, Z. P., Li, Z. F. & Qiao, Y. Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks. Ieee Signal Proc Let 23, 1499-1503, https://doi:10.1109/Lsp.2016.2603342 (2016).

[9] Qin, B. & Li, D. (2020), Identifying Facemask-wearing Condition using Image Super-resolution with ClassificationNetwork to Prevent COVID-19. https://doi.org/10.21203/rs.3.rs-28668/v1

[10] S. V. Militante, B. D. Gerardo, and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IoT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2019, pp. 579-582, https://doi: 10.1109/ECICE47484.2019.8942686.Sammy Sugarcane

[11] S. V. Militante, B. D. Gerardo, and R. P. Medina, "Sugarcane Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IoT, Communication, and Engineering (ECICE), Yunlin, Taiwan, 2019, pp. 575-578, https://doi: 10.1109/ECICE47484.2019.8942690.

[12] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[15] D. C. Cires¸an, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, High-Performance Convolutional Neural Networks for Image Classification," in Proceedings of the twenty-second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 2011.

[16] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, 2012.

[17] D. H. Hubel, and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," The Journal of Physiology, vol. 195, no. 1, p. 215–243, 1968.